## Project Information

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Identifier</td>
<td>PID TBC</td>
</tr>
<tr>
<td>Project Title</td>
<td>UK Research Data (Metadata) Registry Pilot</td>
</tr>
<tr>
<td>Project Hashtag</td>
<td>#jiscrdr</td>
</tr>
<tr>
<td>Start Date</td>
<td>1 October 2013</td>
</tr>
<tr>
<td>End Date</td>
<td>31 March 2014</td>
</tr>
<tr>
<td>Lead Institution</td>
<td>Jisc</td>
</tr>
<tr>
<td>Project Director</td>
<td>Rachel Bruce</td>
</tr>
<tr>
<td>Project Manager</td>
<td>—</td>
</tr>
<tr>
<td>Contact Email</td>
<td>—</td>
</tr>
<tr>
<td>Partner Institutions</td>
<td>Digital Curation Centre (Universities of Edinburgh, Glasgow, Bath); UKDA (University of Essex)</td>
</tr>
<tr>
<td>Project Webpage URL</td>
<td>TBC</td>
</tr>
<tr>
<td>Programme Name</td>
<td>Jisc Capital Programme</td>
</tr>
</tbody>
</table>

## Document Information

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s) and Role(s)</td>
<td>Alex Ball (Metadata Coordinator)</td>
</tr>
<tr>
<td>Date</td>
<td>12 February 2014</td>
</tr>
<tr>
<td>Project Refs</td>
<td>T3.1; D3.1</td>
</tr>
<tr>
<td>Filename</td>
<td>uk-rdr-mapping-v06.pdf</td>
</tr>
<tr>
<td>URL</td>
<td>TBD</td>
</tr>
<tr>
<td>Access</td>
<td>This report is for general dissemination</td>
</tr>
</tbody>
</table>

## Document History

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>29 November 2013</td>
<td>Summary of RIF-CS. First draft of DDI mapping.</td>
</tr>
<tr>
<td>02</td>
<td>8 January 2014</td>
<td>Second draft of DDI mapping. First draft of DataCite mapping.</td>
</tr>
<tr>
<td>04</td>
<td>23 January 2014</td>
<td>Section on tips for contributors.</td>
</tr>
<tr>
<td>05</td>
<td>31 January 2014</td>
<td>Second draft of NERC Discovery (UK GEMINI), EPrints mappings.</td>
</tr>
<tr>
<td>06</td>
<td>12 February 2014</td>
<td>Third draft of DDI mapping.</td>
</tr>
</tbody>
</table>
# Contents

1 Introduction ........................................ 3
   1.1 Typographical conventions .......................... 3

2 RIF-CS ........................................ 4
   2.1 Elements ........................................... 4
   2.2 Controlled vocabularies ................................ 8

3 Internally managed RIF-CS elements ................................ 20

4 Mapping from DDI to RIF-CS ................................ 21
   4.1 Related Objects ........................................ 22

5 Mapping from UK GEMINI 2 to RIF-CS ....................... 24
   5.1 Related Objects ........................................ 26

6 Mapping from DataCite to RIF-CS .......................... 27
   6.1 Related Objects ........................................ 28
   6.2 Related Information .................................... 29

7 Mapping from EPrints to RIF-CS ........................... 31
   7.1 Related Objects ........................................ 32

8 Mapping from OAI-PMH Dublin Core to RIF-CS ............ 34
   8.1 Related Objects ........................................ 35

9 Tips for contributors .................................... 36
   9.1 High priority metadata from harvested records ........... 36

Bibliography ........................................ 40
Introduction

This report documents the mappings created for importing metadata into the pilot UK Research Data Registry implementation. As the priority for the pilot is for populating the registry rather than surfacing content elsewhere, the mappings are all one-way, and convert metadata into the registry’s internal data format, described in Chapter 2.

The target metadata schemas were determined by the capabilities of the project collaborators, i.e. those organizations who volunteered to provide the project with test metadata.

- The UK Data Archive provided records using the DDI standard, version 2.5.
- The NERC Data Catalogue Service provided records using UK GEMINI, version 2.2.
- The University of Edinburgh (using DSpace) provided records using OAI-PMH Dublin Core.
- The University of Glasgow provided records using an adapted EPrints metadata profile.
- The University of Oxford (using DataBank) provided records using ???
- Oxford Brookes University provided records using ???
- The University of Lincoln provided records using ???
- The University of Leeds provided records using an adapted EPrints metadata profile.
- The University of St Andrews provided records using ???
- The University of Southampton provided records using an adapted EPrints metadata profile.
- The University of Hull provided records using ???

1.1 Typographical conventions

In the mappings presented in this report, nesting levels are indicated using angle brackets, while attributes are presented in square brackets. For example, the full path of the type of citation date in a RIF-CS record would be written as registryObjects > registryObject > collection > citationInfo > citationMetadata > date[TYPE], though for brevity the beginning of the path may be omitted.

If the value of an attribute is relevant to identifying an element, or would be fixed for given instance of an element, it is presented alongside the attribute name with an equals sign, e.g. date[TYPE=created].
The ANDS data registry software, used in the initial pilot phase of the UK Research Data Registry project, uses RIF-CS as its internal data model.¹ It is a profile of ISO 2146, tailored to the needs of a collection service registry.

A summary of the RIF-CS standard follows. It is presented as a structured list of elements and their attributes. Indentation in the list corresponds to the nesting of elements, so for example if an element contains two sub-elements, these would be displayed after the element at one greater level of indentation. Attributes immediately follow the elements to which they belong.

Each item in the list has the following structure.

- An expression in square brackets indicates how often the item should occur in a conforming instance of the metadata scheme. An element is represented by two numbers: a minimum and a maximum number of occurrences. An ‘n’ indicates that there is no upper bound to the number of times an element may occur. An attribute is either optional (‘O’) or mandatory (‘M’).

- The item name is given: element names in a *sans serif* typeface, attribute names in *sans serif small capitals*.

- If the item has special formatting rules, this is given in roman type after the name.
  - ‘encoded’ means that the content of the element or the value of the attribute must conform to a particular syntax; where possible the syntax is given in parentheses.
  - ‘controlled’ means that the content of the element or the value of the attribute must be one of a number of pre-defined terms; where possible the authority list of terms is given in a footnote. Definitions of many of these terms are provided in Section 2.2.

- Where the meaning or usage of the item is not clear, additional information is provided in *italics*.

### 2.1 Elements

[1,1] registryObjects

- [0,n] registryObject

  - [M] GROUP – Name identifying the organisation contributing the object’s metadata.

---

key – Unique identifier (used by Registry).

originatingSource – Identifying string or URI of the entity holding the managed version of the registry object metadata.

one of

activity

identifier – May repeat key or supply alternative identifiers.

name

namePart – Can be used for a whole name or for an element of a name.

location – Not used for recording the repository holding a collection, nor for data coverage.

electronic

value: encoded (URI) – URI of collection, activity or party; base URL for HTTP service point; URL of WSDL file.

arg – Argument for an electronic service (not used for activities, collections or parties).

physical

addressPart – May contain full address or a meaningful fragment (e.g. postcode).

spatial – Geographical address information.

relatedObject – Used for other objects in the Registry (c.f. relatedInfo).

key – Unique identifier used by Registry.

Additional information:

2. [http://www.w3.org/TR/xmlschema-2/#dateTime](http://www.w3.org/TR/xmlschema-2/#dateTime)
· · · · · · [1,n] relation
  L [M] TYPE: controlled (see Sections 2.2.18 to 2.2.21)
· · · · · · · · [0,1] description – Required for ‘hasAssociationWith’ relations.
· · · · · · · · [0,1] url: encoded (URL)
· · · · · · [0,n] subject – Keyword indicating field of activity, subject matter of collection, research interest of party.
  L [M] TYPE: controlled¹ – Name of controlled vocabulary used, if any.
· · · · · · · · [0,n] description
  L [M] TYPE: controlled (see Section 2.2.23)
· · · · · · [0,n] coverage
· · · · · · · · [0,n] temporal
· · · · · · · · · · [0,n] date: encoded
  L [M] TYPE: controlled (see Section 2.2.10)
  L [M] DATEFORMAT: controlled (‘W3CDTF’)
· · · · · · · · · · [0,n] text – Used if exact dates are not appropriate.
· · · · · · · · [0,n] spatial – Geographical coverage, e.g. co-ordinates, region information.
  L [M] TYPE: controlled (see Section 2.2.17)
· · · · · · · · [0,n] relatedInfo – Used for objects outside the Registry (c.f. relatedObject).
  L [O] TYPE: controlled (see Section 2.2.24)
· · · · · · · · [1,n] identifier
  L [M] TYPE: controlled (see Section 2.2.25)
· · · · · · [0,n] relation
  L [M] TYPE: controlled (see Sections 2.2.18 to 2.2.21)
· · · · · · · · [0,1] description – Required for ‘hasAssociationWith’ relations.
· · · · · · · · [0,1] url: encoded (URL) – Can be used to record URL implementing a relationship, e.g. for a collection, the URL implementing the related service.
· · · · · · · · [0,1] format
· · · · · · · · · · [1,n] identifier
  L [M] TYPE: controlled (see Section 2.2.25)
· · · · · · · · [0,1] title
· · · · · · · · · · [0,1] notes
· · · · · · · · [0,n] rights
· · · · · · · · · · [0,1] rightsStatement – IPR statement, licence, access rights/constraints statement.
  L [O] RIGHTSURI: encoded (URI) – Used in place of free text content.
· · · · · · · · · · [0,1] licence – Text of legal document.
  L [O] TYPE: controlled (see Section 2.2.26)

3. A term from the Library of Congress’ Source Codes for Vocabularies, Rules, and Schemes (http://www.loc.gov/standards/sourcelist/), or ‘local’ (indicating a controlled vocabulary not in the list, or an uncontrolled vocabulary). The UK registry has defined some additional terms: see Section 2.2.22.
Access Rights – Description of access rights and constraints.

Existence Dates – Refers to existence of registry object.

Start Date: encoded

End Date: encoded

Collection

Dates – (Typically provided after name and before location).

Citation Info – Used to provide a sample bibliographic reference.

Full Citation – Pre-formatted reference.

Citation Metadata – Elements from which to construct a reference.

Identifier

Contributor

Name Part – Can be used for a whole name or for an element of a name.

Title

Version

Publisher

Place Published

Date: encoded (W3C dateTime, UTC)

Url: encoded (URL)

Context – e.g. series title, database name.

Party

As activity plus…
As activity

- · · · service
  - [M] TYPE: controlled (see Section 2.2.5)
  - [O] DATEMODIFIED: encoded (W3C dateTime, UTC) – Refers to the object’s metadata.
  
  As activity plus…

- · · · [0,n] accessPolicy: encoded (URL) – Web-accessible description of service access policies.

### 2.2 Controlled vocabularies

The following represent the controlled terms defined for use in RIF-CS 2.5, and are reproduced here for convenience. The headings indicate where, within a registryObject element, the vocabularies are used (vertical bars indicate alternatives at that part of the path hierarchy). They are presented in the order in which they would be first encountered in an XML record, if the element order in the previous section were used.

Substantive deviations from the published definitions have been marked in italic. These are intended as corrections or clarifications rather than local variations.

#### 2.2.1 originatingSource[type]

authoritative The source holds the authoritative version of the metadata about the registry object.

#### 2.2.2 activity[type]

award Something given to recognize excellence in a certain field.
course Education imparted in a series of lessons or meetings.
event Something that happens at a particular place or time as an organized activity with participants or an audience.
program System of activities intended to meet a public need.
project Piece of work that is undertaken or attempted, with a start and end date and defined objectives.

#### 2.2.3 collection[type]

catalogueOrIndex Collection of resource descriptions describing the content of one or more repositories or collective works.
collection Compiled content created as separate and independent works and assembled into a collective whole for distribution and use.
registry Collection of registry objects compiled to support the business of a given community.
repository Collection of physical or digital objects compiled for information and documentation purposes and/or for storage and safekeeping.
dataset Collection of physical or digital objects generated by research activities.

---

2.2.4 party[type]

group One or more persons acting as a family, group, association, partnership or corporation.
person Human being or identity assumed by one or more human beings.
administrativePosition A kind of party where the position, name and contact information are present but the identity of the party filling the role is not specified.

2.2.5 service[type]

create Instrument.
generate Simulator.
report Visualisation, summary.
annotate Collection of comments, reviews, ratings.
transform Analysis, conversion.
assemble Aggregation.
harvest-oaipmh OAI-PMH Harvest.
search-http Search service over HTTP.
search-opensearch OpenSearch search.
search-sru SRU search.
search-srw SRW search.
search-z3950 Z39.50 search.
syndicate-atom ATOM syndication.
syndicate-rss RSS feed.

2.2.6 activity | collection | party | service > identifier[type]

abn Australian Business Number.
arc Australian Research Council identifier.
ark ARK Persistent Identifier Scheme.
doi Digital Object Identifier.
handle Handle System Identifier.
infouri 'Info' URI scheme.
isil International Standard Identifier for Libraries.
local Identifier unique within a local context.
hnhmrc National Health and Medical Research Council identifier.
orcid Open Researcher and Contributor Identifier.
purl Persistent Uniform Resource Locator.
uri Uniform Resource Identifier.
2.2.7 activity | collection | party | service > name[type]

**primary** Official name of the registry object.

**abbreviated** Shortened form of, or acronym for, the official name.

**alternative** Any other form of name used now or in the past as a substitute or alternative for the official name.

2.2.8 party > name > namePart[type], collection > citationInfo > citation-Metadata > contributor > namePart[type]

**family** Last name or surname.

**given** Forename or given or Christian name.

**suffix** Honours, awards, qualifications and other identifiers conferred.

**title** Word or phrase indicative of rank, office, nobility, honour, etc., or a term of address associated with a person.

**superior** Part of a name that describes a party (group) that contains one or more integral subordinate parties (sub-groups or sub-units).

**subordinate** Part of a name that describes a party (group) that is an integral sub-group or sub-unit of a superior party (group).

2.2.9 collection > dates[type]

**dc.available** Date (often a range) that the resource became or will become available.

**dc.created** Date of creation of the resource.

**dc.dateAccepted** Date of acceptance of the resource.

**dc.dateSubmitted** Date of submission of the resource.

**dc.issued** Date of formal issuance (e.g. publication) of the resource.

**dc.valid** Date (often a range) of validity of a resource.

2.2.10 collection > dates > date[type], activity | collection | party | service > coverage > temporal > date[type]

**dateFrom** Start date for the applicability of the date, or a temporal coverage period.

**dateTo** End date for the applicability of the date, or a temporal coverage period.

2.2.11 collection > dates > date[dateFormat], activity | collection | party | service > coverage > temporal > date[dateFormat]

**W3CDTF** W3C Date Time Format.

2.2.12 activity | collection | party | service > location > address > electronic[type]

**Service**

**wsdl** Web Service Definition Language.
All

e-mail  String used to receive messages by means of a computer network.
other  Other electronic address.
url  Uniform Resource Locator.

2.2.13 service > location > address > electronic > arg[type]

string Indicates the value of an argument is a plain text string
object indicates the value of an argument is an object, most likely in serialized form

2.2.14 service > location > address > electronic > arg[use]

inline Indicates the argument forms part of the base URL.
keyValue Indicates the argument is passed using key = value pairings in the query component of a URL.

2.2.15 activity | collection | party | service > location > address > physical[type]

streetAddress Address where an entity is physically located.
postalAddress Address where mail for an entity should be sent.

2.2.16 activity | collection | party | service > location > address > physical > addressPart[type]

addressLine An address part that is a separate line of a structured address.
text A single address part that contains the whole address in unstructured form.
telephoneNumber An address part that contains a telephone number, including a mobile telephone number.
faxNumber An address part that contains a fax (facsimile) number.

2.2.17 activity | collection | party | service > location | coverage > spatial[type]

gmlKmlPolyCoords A set of KML long/lat co-ordinates derived from GML (OpenGIS Geography Markup Language) defining a polygon as described by the KML co-ordinates element but without the altitude component.
gpx GPS Exchange Format.
iso31661 ISO 3166-1 Codes for the representation of names of countries and their subdivisions - Part 1: Country codes.
iso31662 Codes for the representation of names of countries and their subdivisions - Part 2: Country subdivision codes.
iso31663 ISO 3166-3 Codes for country names which have been deleted from ISO 3166-1 since its first publication in 1974.
iso19139dcmiBox DCMI Box notation derived from bounding box metadata conformant with the ISO 19139 schema.
kmlPolyCoords  A set of KML (Keyhole Markup Language) long/lat co-ordinates defining a polygon as described by the KML coordinates element.
dcmiPoint  spatial location information specified in DCMI Point notation.
text  Free-text representation of spatial location.

2.2.18  activity > relatedObject | relatedInfo > relation[type]

To activity

hasPart  Contains the related activity.
isPartOf  Is contained in the related activity.

To collection

hasOutput  Delivers materials in the related collection.

To party

hasParticipant  Is undertaken by the related party.
hasPrincipalInvestigator  Is researched by the related party.
isFundedBy  Receives monetary or in-kind aid from the related program.
isManagedBy  Is organised and/or delivered by the related party.
isOwnedBy  Legally belongs to the related party.

To any

hasAssociationWith  Has an unspecified relationship with the related object.

2.2.19  collection > relatedObject | relatedInfo > relation[type]

To activity

isOutputOf  Is a product of the related activity.

To collection

describes  Is a catalogue for, or index of, of items in the related collection.
hasPart  Contains the related collection.
isDescribedBy  Is catalogued or indexed by the related catalogue or index.
isLocatedIn  Is held in the related repository.
isLocationFor  Is the repository where the related collection is held.
isPartOf  Is contained within the related collection.
isDerivedFrom  Collection is derived from the related collection, e.g. through analysis.
hasDerivedCollection  The related collection is derived from the collection, e.g. through analysis.
To party

hasCollector Has been aggregated by the related party.
hasPrincipalInvestigator Is researched by the related party.
isManagedBy Is maintained and made accessible by the related party.
isOwnedBy Legally belongs to the related party.
isEnrichedBy Additional value provided to a collection by a party.

To publication

isCitedBy Indicates that B includes A in a citation.
isReferencedBy Indicates A is used as a source of information by B.
isDocumentedBy Indicates B is documentation about/explaining A.
isSupplementedBy Indicates that A is a supplement to B.
isSupplementTo Indicates that B is a supplement to A.
isReviewedBy The cited entity presents statements, ideas or conclusions that are reviewed by the citing entity.
isSupportedBy The cited entity receives intellectual or factual support from the citing entity.

To service

supports Can be contributed to, accessed or used through the related service.
isAvailableThrough Specialisation of ‘supports’ type, for Harvest, Search and Syndicate.
isProducedBy Specialisation of ‘supports’ type, for Create, Generate and Assemble.
isPresentedBy Specialisation of ‘supports’ type, for Report.
hasValueAddedBy Specialisation of ‘supports’ type, for Annotate.
isOperatedOnBy Specialisation of ‘supports’ type, for Transform.

To any

hasAssociationWith Has an unspecified relationship with the related object.

2.2.20 party > relatedObject | relatedInfo > relation[type]

To activity

isFundedBy Receives monetary or in-kind aid from the related program.
isFunderOf Provides monetary or in-kind aid to the related activity.
isParticipantIn Is enrolled in the related activity.
isPrincipalInvestigatorOf Is a researcher involved in the related activity.
To collection

enriches Provides additional value to the related collection.
isCollectorOf Has aggregated the related collection.
isManagerOf Administers the related collection.
isPrincipalInvestigatorOf Is a researcher involved in the related collection.

To party

hasMember Is a group that has enrolled the related party.
hasPart Is a group that contains the related group.
isFundedBy Receives monetary or in-kind aid from the related party.
isFunderOf Provides monetary or in-kind aid to the related party.
isManagedBy Is overseen by the related party.
isManagerOf Oversees the related party.
isMemberOf Is enrolled in the related group.
isOwnedBy Legally belongs to the related party.
isPartOf Is a group that is contained in the related group.

To any

hasAssociationWith Has an unspecified relationship with the related object.
isOwnerOf Legally possesses the related activity, collection, service or group.

2.2.21 service > relatedObject | relatedInfo > relation[type]

To collection

isSupportedBy Enables contribution and access to and use of the related collection.
makesAvailable Specialisation of ‘isSupportedBy’ type, for Harvest, Search and Syndicate.
produces Specialisation of ‘isSupportedBy’ type, for Create, Generate and Assemble.
presents Specialisation of ‘isSupportedBy’ type, for Report.
operatesOn Specialisation of ‘isSupportedBy’ type, for Transform.
addsValueTo Specialisation of ‘isSupportedBy’ type, for Annotate.

To party

isManagedBy Is overseen by the related party.
isOwnedBy Legally belongs to the related party.

To service

hasPart Contains the related service.
isPartOf Is contained in the related service.
To any

`hasAssociationWith` Has an unspecified relationship with the related `object`.

### 2.2.22 activity | collection | party | service > subject[type]

The following are additional controlled terms used by the UK registry but not recognized in the RIF-CS standard.

- `gemet` General Environmental Multi-Lingual Thesaurus.⁵
- `hasset` Humanities and Social Science Electronic Thesaurus.⁶
- `ipsv2` Integrated Public Sector Vocabulary, version 2.0.⁷
- `iso19115topic` ISO 19115 Topic Category Code; see for example UK GEMINI⁸ for a list.
- `jacs3` Joint Academic Coding System (JACS), version 3.0.⁹
- `rcukrc` Research Councils UK Research Classifications.¹⁰
- `ukdasc` UK Data Archive Subject Categories.

### 2.2.23 activity | collection | party | service > description[type]

**Collection**

- `significanceStatement` A statement describing the significance of a collection within its domain or context.
- `lineage` Text describing the collection lineage

**Party**

- `researchAreas` Text describing a contributor organisation’s distinctive research portfolio and research strengths.
- `researchDataProfile` Text describing and highlighting the research data (and related parties, projects and services) whose description the organisation has contributed to Research Data Australia.
- `researchSupport` Text describing specific data-related support services offered by the contributor organisation such as archives, repositories, data centres, metadata stores, high performance computing facilities, data-intensive instruments, e-research support centres, data management support services, etc.

---

**Service**

deliveryMethod  Information about how the service is delivered. Should be one of: web-service, software, offline, workflow.

**All**

brief  Short account for selection purposes.

full  Full account.

logo  Symbol used as an identifying mark.

note  A brief informational message, not object metadata, to notify the record consumer of some important aspect regarding the object or its metadata.

**2.2.24 activity | collection | party | service > relatedInfo[type]**

activity  An undertaking or process related to the creation, update, or maintenance of a collection.

collection  An aggregation of physical and/or digital resources which has meaning in a research context.

dataQualityInformation  Data quality statements or summaries of data quality issues affecting the data.

metadata  An alternative metadata format for the Object. This is most likely to be a discipline or system-specific format, e.g. NetCDF or ANZLIC.

party  A person, group or role related to the creation, update, or maintenance of a collection, to an activity, or to the provision of a service.

publication  Any formally published document, whether available in digital or online form or not.

reuseInformation  Information that supports reuse of data, such as data definitions, instrument calibration or settings, units of measurement, sample descriptions, experimental parameters, methodology, data analysis techniques, or data derivation rules.

service  A system (analogue or digital) that provides one or more functions of value to an end user.

website  Any publicly accessible web location containing information related to the collection, activity, party or service.

**2.2.25 activity | collection | party | service > relatedInfo > identifier[type], activity | collection | party | service > relatedInfo > format > identifier[type]**

abn  Australian Business Number.

arc  Australian Research Council identifier.

ark  ARK Persistent Identifier Scheme.


doi  Digital Object Identifier.
ean13  International Article Number.
eissn  Electronic International Standard Serial Number.
handle  Handle System Identifier.
infouri  'Info' URI scheme.
isbn  International Standard Book Number.
isil  International Standard Identifier for Libraries.
issn  International Standard Serial Number.
liissn  Linking International Standard Serial Number.
local  Identifier unique within a local context.
mediaType  The Media Type (MIME type) of the information. Values should be taken from IANA Media Type assignment listing. You may choose to use application/x-name if it is well known within the relevant discipline.
hnmrc  National Health and Medical Research Council identifier.
orcid  Open Researcher and Contributor Identifier.
purl  Persistent Uniform Resource Locator.
researcherID  Thomson Reuters ResearcherID.
upc  Universal Product Code.
uri  Uniform Resource Identifier.
urn  Uniform Resource Name.

2.2.26  activity | collection | party | service > rights > licence[type]

CC-BY  Creative Commons Attribution.
CC-BY-SA  Creative Commons Attribution-ShareAlike.
CC-BY-ND  Creative Commons Attribution-NoDerivs.
CC-BY-NC  Creative Commons Attribution-NonCommercial.
CC-BY-NC-SA  Creative Commons Attribution-NonCommercial-ShareAlike.
CC-BY-NC-ND  Creative Commons Attribution-NonCommercial-NoDerivs.
GPL  GNU General Public Licence.
AusGoalRestrictive  AusGoal Restrictive Licence.
NoLicence  No licence.
Unknown/Other  Unknown licence or custom licence defined by the rights holder.

2.2.27  collection > citationInfo > fullCitation[style]

Harvard  Generic author-date style.
APA  American Psychological Association.
MLA  Modern Language Association of America.
IEEE Institute of Electrical and Electronic Engineers.
CSE Council of Science Editors.
Chicago Chicago Manual of Style.
AMA American Medical Association.
AGPS-AGIMO Australian Style Manual.
AGLC Australian Guide to Legal Citation.
ACS American Chemical Society.
DataCite DataCite Consortium.

2.2.28  collection > citationInfo > citationMetadata > identifier[type]

ark ARK Persistent Identifier Scheme.
doi Digital Object Identifier.
ean13 International Article Number.
eissn Electronic International Standard Serial Number.
handle Handle System Identifier.
infouri ‘Info’ URI scheme
isbn International Standard Book Number.
issn International Standard Serial Number.
lissn Linked International Standard Serial Number.
local Identifier unique within a local context.
purl Persistent Uniform Resource Locator.
upc Universal Product Code.
uri Uniform Resource Identifier.
urn Uniform Resource Name.

2.2.29  collection > citationInfo > citationMetadata > date[type]

publicationDate The date when the data was or will be made publicly available.
available The date the resource is made publicly available. May be a range, or indicate the end of an embargo period
created The date the resource itself was put together; this could be a date range or a single date for a final component, e.g. the finalised file with all of the data
date Any relevant date not otherwise specified
dateAccepted The date that the publisher accepted the resource into their system
dateSubmitted The date the creator submits the resource to the publisher. This could be different from dateAccepted if the publisher then applies a selection process
endPublicationDate  Use when publicationDate is a range.

issued  The date that the resource is published or distributed, e.g. to a data centre.

modified  The date of the last update to the resource, when the resource is being added to.
          (Equivalent to DataCite ‘Updated’.)

startPublicationDate  Use when publicationDate is a range.

valid  The date or date range during which the dataset or resources are accurate.
Internally managed RIF-CS elements

Several of the elements in a RIF-CS record are managed internally by the Registry and are not populated by harvested metadata. These are listed below.

*registryObject*[GROUP] This is set to the name of the organization from which the metadata has been harvested.

*registryObject > key* This is the Registry’s internal unique identifier for the object. It may be based on an existing identifier for the object or newly generated.

*registryObject > originatingSource* This is set to an identifier or URI for the repository that originally supplied the metadata. This is assumed to be the organization from which the metadata has been harvested unless more accurate information can be inferred.

*registryObject > collection*[TYPE]* This is ‘dataset’ for harvested dataset records.

*registryObject > collection*[DATEMODIFIED]* This is the date on which the Registry record was last modified (or created).

*registryObject > collection > existenceDates > startDate* This is the date on which the record was first harvested.

*registryObject > collection > existenceDates > endDate* This would be the date on which the record was withdrawn.
## Mapping from DDI to RIF-CS

The following table provides a mapping to populate a RIF-CS Collection record from a UKDA DDI record. The value of the UKDA ID is recorded in the DDI record at `codeBook > stdyDscr > citation > titlStmt > IDNo[AGENCY=UKDA]`.

<table>
<thead>
<tr>
<th>RIF-CS 1.5 element</th>
<th>Source (using DDI 2.5 record)</th>
</tr>
</thead>
<tbody>
<tr>
<td>collection[dateAccessioned]</td>
<td><code>codeBook &gt; stdyDscr &gt; citation &gt; distStmt &gt; depDate[date]</code></td>
</tr>
<tr>
<td>identifier[type=doi]</td>
<td><code>codeBook &gt; stdyDscr &gt; citation &gt; titlStmt &gt; IDNo[agency=datacite]</code></td>
</tr>
<tr>
<td>identifier[type=local]</td>
<td><code>codeBook &gt; stdyDscr &gt; citation &gt; titlStmt &gt; IDNo[(agency):content]</code></td>
</tr>
<tr>
<td>name[type=primary] &gt; namePart</td>
<td><code>codeBook &gt; stdyDscr &gt; citation &gt; titlStmt &gt; titl</code></td>
</tr>
<tr>
<td>name[type=alternative] &gt; namePart</td>
<td><code>codeBook &gt; stdyDscr &gt; citation &gt; titlStmt &gt; altTitl</code></td>
</tr>
<tr>
<td>dates[type=dc.available, dc.issued] &gt; date[type=dateFrom]</td>
<td><code>codeBook &gt; stdyDscr &gt; citation &gt; distStmt &gt; distDate</code></td>
</tr>
<tr>
<td>dates[type=dc.dateSubmitted] &gt; date[type=dateFrom]</td>
<td><code>codeBook &gt; stdyDscr &gt; citation &gt; distStmt &gt; depDate[date]</code></td>
</tr>
<tr>
<td>location &gt; address &gt; electronic[type=url] &gt; value</td>
<td><code>codeBook &gt; stdyDscr &gt; citation &gt; holdings[URI]</code></td>
</tr>
<tr>
<td>subject[type=hasset]</td>
<td><code>codeBook &gt; stdyDscr &gt; stdyInfo &gt; subject &gt; keyword[vocab=S]</code></td>
</tr>
<tr>
<td>subject[termIdentifier]</td>
<td><code>codeBook &gt; stdyDscr &gt; stdyInfo &gt; subject &gt; keyword[vocab=S; vocabURI]</code></td>
</tr>
<tr>
<td>subject[type=ukdasc]</td>
<td><code>codeBook &gt; stdyDscr &gt; stdyInfo &gt; subject &gt; topClas</code></td>
</tr>
<tr>
<td>description[type=full]</td>
<td><code>codeBook &gt; stdyDscr &gt; stdyInfo &gt; abstract</code></td>
</tr>
<tr>
<td>coverage &gt; temporal &gt; date[type=dateFrom]</td>
<td><code>codeBook &gt; stdyDscr &gt; stdyInfo &gt; sumDscr &gt; collDate[event=start, single; date], timePrd[event=start, single; date]</code></td>
</tr>
<tr>
<td>coverage &gt; temporal &gt; date[type=dateTo]</td>
<td><code>codeBook &gt; stdyDscr &gt; stdyInfo &gt; sumDscr &gt; collDate[event=end; date], timePrd[event=end; date]</code></td>
</tr>
<tr>
<td>coverage &gt; spatial[type=text]</td>
<td><code>codeBook &gt; stdyDscr &gt; stdyInfo &gt; sumDscr &gt; geogCover, geogUnit, nation; codeBook &gt; stdyDscr &gt; stdyInfo &gt; subject &gt; keyword[vocab=G]</code></td>
</tr>
<tr>
<td>relatedInfo[type=metadata] &gt; identifier[type=uri]</td>
<td><code>'http://esds.ac.uk/DDI25/’ + UKDA ID + ‘.xml’</code></td>
</tr>
</tbody>
</table>
4.1 Related Objects

- A series is represented in the Registry as a Collection of type ‘collection’. Each member of the series is listed in the series object as a Related Object with Relation type ‘hasPart’. Within each member object, the series object is included as a Related Object with Relation type ‘isPartOf’.

- The holding repository (i.e. UKDA) is represented in the Registry as a Collection of type ‘repository’. Each dataset object includes the repository object as a Related Object with Relation type ‘isLocatedIn’. The repository object lists each dataset object as a Related Object with Relation type ‘isLocationFor’.

- Individuals and corporate entities contributing to the dataset are represented in the Registry as Parties. The dataset object lists each contributor as a Related Object with Relation type ‘hasPrincipalInvestigator’. Each contributor object includes the dataset object as a Related Object with Relation type ‘isPrincipalInvestigatorOf’.

1. All records should have a version statement, but in case not the crosswalk falls back to ‘1’.
• Funding information (recorded in the UKDA DDI records under `codeBook > stdyDscr > citation > prodStmt > fundAg`) would most naturally be represented via an Activity object. The chain of relationships would be

Collection (dataset) isOutputOf Activity (study) isFundedBy Party (funder),

and conversely,

Party isFunderOf Activity hasOutput Collection.

It is, however, currently unclear how to populate the Activity record.
The following table provides a mapping to populate a RIF-CS Collection record from one held by the NERC Data Catalogue Service. The mapping is based on UK GEMINI,¹ which is a profile of ISO 19115.

In considering whether this mapping might be suitable for other contexts, it may be helpful to consider MEDIN Discovery² to be a special case of NERC Discovery³ – the profile originally used by the Data Catalogue Service – itself a special case of UK GEMINI, which is a special case of INSPIRE.⁴ All of these are ISO 19115 profiles.

For brevity the element numbers and names from UK GEMINI are used, rather than the full XML paths from the ISO 19115 standard.

<table>
<thead>
<tr>
<th>RIF-CS 1.5 element</th>
<th>Source (using NERC record)</th>
</tr>
</thead>
<tbody>
<tr>
<td>originatingSource</td>
<td>35 – Metadata point of contact &gt; Organisation name, Email address (translated to identifier or URI)</td>
</tr>
<tr>
<td>identifier</td>
<td>36 – Unique resource identifier &gt; Code</td>
</tr>
<tr>
<td>identifier[type]</td>
<td>36 – Unique resource identifier &gt; Code space</td>
</tr>
<tr>
<td>name[type=primary]</td>
<td>1 – Title</td>
</tr>
<tr>
<td>name[type=alternative]</td>
<td>2 – Alternative title</td>
</tr>
<tr>
<td>dates[type=dc.created]</td>
<td>8 – Dataset reference date &gt; Date type = creation, Date</td>
</tr>
<tr>
<td>dates[type=dc.issued]</td>
<td>8 – Dataset reference date &gt; Date type = publication, Date</td>
</tr>
<tr>
<td>location &gt; address &gt;</td>
<td>36 – Unique resource identifier if resolvable, otherwise 19 – Resource locator &gt; Resource locator URL⁶</td>
</tr>
<tr>
<td>electronic[type=url]</td>
<td>value</td>
</tr>
</tbody>
</table>

¹. AGI, UK GEMINI.
⁵. A single date value in the GEMINI record would be represented by date[type=dateFrom] in the RIF-CS record. A date range in the GEMINI record, such as ‘2004-03-02/2005-06-02’, would be represented by date[type=dateFrom] (‘2004-03-02’) and date[type=dateTo] (‘2005-06-02’) in the RIF-CS record.
⁶. Where multiple Resource locator URLs are provided, one will be selected using the value of Resource locator function. The order of preference is download, order, offlineAccess, information, search, no specified function.
<table>
<thead>
<tr>
<th>RIF-CS 1.5 element</th>
<th>Source (using NERC record)</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject[type=iso19115topic]</td>
<td>5 – Topic category (ISO 19115 code list)</td>
</tr>
<tr>
<td>subject</td>
<td>6 – Keyword &gt; Keyword value</td>
</tr>
<tr>
<td>subject[type]</td>
<td>6 – Keyword &gt; Originating controlled vocabulary</td>
</tr>
<tr>
<td>description[type=full]</td>
<td>4 – Abstract</td>
</tr>
<tr>
<td>description[type=lineage]</td>
<td>10 – Lineage</td>
</tr>
<tr>
<td>coverage &gt; temporal &gt; date⁵</td>
<td>7 – Temporal extent</td>
</tr>
<tr>
<td>coverage &gt; spatial[type=iso19139dcmiBox]</td>
<td>44 – Bounding box</td>
</tr>
<tr>
<td>coverage &gt; spatial</td>
<td>5 – Extent &gt; Extent name</td>
</tr>
<tr>
<td>coverage &gt; spatial[type]</td>
<td>5 – Extent &gt; Originating controlled vocabulary &gt; Thesaurus name</td>
</tr>
<tr>
<td>relatedInfo[type=publication, website] &gt; identifier, identifier[type], title</td>
<td>Inferred from text processing 27 – Additional information source and 19 – Resource locator (with function ‘information’) if possible</td>
</tr>
<tr>
<td>relatedInfo &gt; relation[type]</td>
<td>For publications, ‘isDocumentedBy’; for websites, ‘hasAssociationWith’</td>
</tr>
<tr>
<td>relatedInfo[type=website] &gt; relation[type=hasAssociationWith] &gt; description</td>
<td>‘Has additional information’</td>
</tr>
<tr>
<td>relatedInfo[type=metadata] &gt; identifier[type=uri]</td>
<td>URL of NERC DCS record⁷</td>
</tr>
<tr>
<td>rights &gt; accessRights</td>
<td>25 – Limitations on Public Access, 26 – Use constraints</td>
</tr>
<tr>
<td>citationInfo &gt; citationMetadata &gt; identifier</td>
<td>36 – Unique resource identifier &gt; Code</td>
</tr>
<tr>
<td>citationInfo &gt; citationMetadata &gt; identifier[type]</td>
<td>36 – Unique resource identifier &gt; Code space</td>
</tr>
<tr>
<td>citationInfo &gt; citationMetadata &gt; contributor &gt; namePart</td>
<td>23 – Responsible organisation &gt; Individual name⁸</td>
</tr>
<tr>
<td>citationInfo &gt; citationMetadata &gt; title</td>
<td>1 – Title</td>
</tr>
<tr>
<td>citationInfo &gt; citationMetadata &gt; publisher</td>
<td>23 – Responsible organisation &gt; Organisation name⁹</td>
</tr>
</tbody>
</table>

7. This is an OGC CSW request with the following form: http://(CSW host)/geonetwork/srv/en/csw?SERVICE=CSW&REQUEST=GetRecordById&ID=〈file identifier〉&elementSetName=full&OutputSchema=http://www.isotc211.org/2005/gmd

8. Responsible party role is used to filter the responsible parties provided. Only one role is used, and they are checked in this order: author, originator, principalInvestigator, owner. If the author role is used, the order will be preserved using the RIF-CS SEQ attribute.

9. Responsible party role is used to filter the responsible parties provided. Only one role is used, and they are checked in this order: publisher, distributor, resource provider. Note that the distributor may be recorded in the ISO 19115 element distributionInfo > … > distributorContact instead of identificationInfo > … > pointOfContact.
5.1 Related Objects

- A series is represented in the Registry as a Collection of type ‘collection’. Each member of the series is listed in the series object as a Related Object with Relation type ‘hasPart’. Within each member object, the series object is included as a Related Object with Relation type ‘isPartOf’.

- Individuals and corporate entities contributing to the dataset are represented in the Registry as Parties. The following table shows the mapping from UK GEMINI’s controlled vocabulary of roles to those used by RIF-CS for types of relationship (note that ‘hasPrincipalInvestigator’ has a broader definition than the name suggests). Entities with other roles are not recorded in the Registry.

<table>
<thead>
<tr>
<th>RIF-CS 1.5 values</th>
<th>UK GEMINI values</th>
</tr>
</thead>
<tbody>
<tr>
<td>isManagedBy/isManagerOf</td>
<td>Custodian</td>
</tr>
<tr>
<td>isOwnedBy/isOwnerOf</td>
<td>Owner</td>
</tr>
<tr>
<td>hasPrincipalInvestigator/isPrincipalInvestigatorOf</td>
<td>Originator</td>
</tr>
<tr>
<td>hasPrincipalInvestigator/isPrincipalInvestigatorOf</td>
<td>Principal Investigator</td>
</tr>
<tr>
<td>isEnrichedBy/enriches</td>
<td>Processor</td>
</tr>
<tr>
<td>hasPrincipalInvestigator/isPrincipalInvestigatorOf</td>
<td>Author</td>
</tr>
</tbody>
</table>

10. Where multiple Resource locator URLs are provided, one will be selected using the value of Resource locator function. The order of preference is download, order, offlineAccess, information, search, no specified function.
The following table provides a mapping to populate a RIF-CS Collection record from a DataCite record.¹

<table>
<thead>
<tr>
<th>RIF-CS 1.5 element</th>
<th>Source (using DataCite record)</th>
</tr>
</thead>
<tbody>
<tr>
<td>collection[dateAccessioned]</td>
<td>date[type=Accepted]</td>
</tr>
<tr>
<td>identifier[type=doi]</td>
<td>Identifier[identifierType=DOI]</td>
</tr>
<tr>
<td>identifier²</td>
<td>AlternativeIdentifier</td>
</tr>
<tr>
<td>name[type=primary] &gt; namePart</td>
<td>Title</td>
</tr>
<tr>
<td>name[type=alternative] &gt; namePart</td>
<td>Title[type=AlternativeTitle]</td>
</tr>
<tr>
<td>dates[type=dc.available] &gt; date³</td>
<td>Date[type=Available]</td>
</tr>
<tr>
<td>dates[type=dc.created] &gt; date³</td>
<td>Date[type=Created]</td>
</tr>
<tr>
<td>dates[type=dc.dateAccepted] &gt; date³</td>
<td>Date[type=Accepted]</td>
</tr>
<tr>
<td>dates[type=dc.dateSubmitted] &gt; date³</td>
<td>Date[type=Submitted]</td>
</tr>
<tr>
<td>dates[type=dc.issued] &gt; date³</td>
<td>Date[type=Issued]</td>
</tr>
<tr>
<td>dates[type=dc.valid] &gt; date³</td>
<td>Date[type=Valid]</td>
</tr>
<tr>
<td>location &gt; address &gt; electronic[type=url] &gt; value</td>
<td>‘<a href="http://dx.doi.org/%E2%80%99">http://dx.doi.org/’</a> + Identifier[identifierType=DOI]</td>
</tr>
<tr>
<td>subject⁴</td>
<td>Subject</td>
</tr>
<tr>
<td>description[type=full]</td>
<td>Description[type=Abstract]</td>
</tr>
<tr>
<td>description[type=lineage]</td>
<td>Description[type=Methods]</td>
</tr>
<tr>
<td>coverage &gt; spatial[type=iso19139dcmibox]</td>
<td>GeoLocation[type=geoLocationBox]</td>
</tr>
<tr>
<td>coverage &gt; spatial[type=dcmiPoint]</td>
<td>GeoLocation[type=geoLocationPoint]</td>
</tr>
<tr>
<td>coverage &gt; spatial[type=text]</td>
<td>GeoLocation[type=geoLocationPlace]</td>
</tr>
</tbody>
</table>

¹. DataCite Metadata Working Group, *DataCite Metadata Schema for the Publication and Citation of Research Data*, version 3.0 (DataCite Consortium, 2013), DOI: 10.5438/0008.

². DataCite does not use a controlled vocabulary for alternateIdentifierType, so some parsing would be needed to ensure that identifier types known to RIF-CS are recognised. The RIF-CS type ‘local’ is used as fallback for unrecognised types.

³. A single date value in the DataCite record would be represented by date[type=dateFrom] in the RIF-CS record. A date range in the DataCite record, such as ‘2004-03-02/2005-06-02’, would be represented by date[type=dateFrom] (‘2004-03-02’) and date[type=dateTo] (‘2005-06-02’) in the RIF-CS record.

⁴. Some interpretation would be needed to translate values of DataCite’s subjectScheme or schemeURI (e.g. ‘http://id.loc.gov/authorities/subjects’) into terms from the controlled vocabularies used by RIF-CS (e.g. ‘lcsh’). The RIF-CS type ‘local’ is used as fallback for unrecognised schemes.
## 6.1 Related Objects

- The holding repository (i.e. DataCite Publisher) is represented in the Registry as a Collection of type ‘repository’. Each dataset object includes the repository object as a Related Object with Relation type ‘isLocatedIn’. The repository object lists each dataset object as a Related Object with Relation type ‘isLocationFor’.

- Individuals and corporate entities contributing to the dataset are represented in the Registry as Parties. The dataset object lists each DataCite Author and Contributor (with contributorTypes ‘DataCollector’, ‘ProjectLeader’, or ‘WorkPackageLeader’) as a Related Object with Relation type ‘hasPrincipalInvestigator’. Each contributor object includes the dataset object as a Related Object with Relation type ‘isPrincipalInvestigatorOf’. (Note that RIF-CS defines these relations as involving a researcher in the collection: it does not specify that the researcher has to be an official PI.)
• Funding information (recorded in DataCite records using Contributor elements with contributorType ‘Funder’) would most naturally be represented via an Activity object. The chain of relationships would be

Collection (dataset) isOutputOf Activity (study) isFundedBy Party (funder),
and conversely,

Party isFunderOf Activity hasOutput Collection.

It is, however, currently unclear how to populate the Activity record.

### 6.2 Related Information

The following table shows the mapping from DataCite’s controlled vocabulary to that used by RIF-CS for related object identifiers.

<table>
<thead>
<tr>
<th>RIF-CS 1.5</th>
<th>DataCite</th>
<th>RIF-CS 1.5</th>
<th>DataCite</th>
<th>RIF-CS 1.5</th>
<th>DataCite</th>
</tr>
</thead>
<tbody>
<tr>
<td>ark</td>
<td>ARK</td>
<td>isbn</td>
<td>ISBN</td>
<td>local</td>
<td>PMID</td>
</tr>
<tr>
<td>doi</td>
<td>DOI</td>
<td>issn</td>
<td>ISSN</td>
<td>purl</td>
<td>PURL</td>
</tr>
<tr>
<td>ean13</td>
<td>EAN13</td>
<td>istc</td>
<td>ISTC</td>
<td>upc</td>
<td>UPC</td>
</tr>
<tr>
<td>eissn</td>
<td>EISSN</td>
<td>lissn</td>
<td>LISSN</td>
<td>uri</td>
<td>URL</td>
</tr>
<tr>
<td>handle</td>
<td>Handle</td>
<td>urn</td>
<td>LSID</td>
<td>urn</td>
<td>URN</td>
</tr>
</tbody>
</table>

The following table shows the mapping from DataCite’s controlled vocabulary for types of relationship, to those used by RIF-CS for types of related object and types of relationship.

<table>
<thead>
<tr>
<th>RIF-CS 1.5 values</th>
<th>DataCite values</th>
</tr>
</thead>
<tbody>
<tr>
<td>publication / isCitedBy</td>
<td>IsCitedBy</td>
</tr>
<tr>
<td>publication / isSupplementedBy</td>
<td>IsSupplementedBy</td>
</tr>
<tr>
<td>publication / isSupplementTo</td>
<td>IsSupplementTo</td>
</tr>
<tr>
<td>collection / isPartOf</td>
<td>IsPartOf</td>
</tr>
<tr>
<td>collection / hasPart</td>
<td>HasPart</td>
</tr>
<tr>
<td>publication / isReferencedBy</td>
<td>IsReferencedBy</td>
</tr>
<tr>
<td>publication / isDocumentedBy</td>
<td>IsDocumentedBy</td>
</tr>
<tr>
<td>collection / isDerivedFrom</td>
<td>isCompiledBy</td>
</tr>
<tr>
<td>collection / hasDerivedCollection</td>
<td>Compiles</td>
</tr>
</tbody>
</table>

All other DataCite relationships should be represented in RIF-CS as relations of type ‘hasAssociationWith’, with the text of the DataCite term put into normal case (e.g. ‘Has metadata’) and added to the RIF-CS relatedInfo > relation > description element.

The DataCite relationships ‘isContinuedBy’, ‘Continues’, ‘isMetadataFor’ ‘isNewVersionOf’, ‘isPreviousVersionOf’, ‘Documents’, ‘isVariantFormOf’, ‘isOriginalFormOf’ and ‘isIdenticalTo’ imply a relatedInfo type of ‘collection’. The relationships ‘Cites’ and ‘References’ imply a relatedInfo type of ‘publication’.

Note that the DataCite relationship ‘HasMetadata’ could be with metadata (i.e. an alternative, full metadata record for the dataset in question), reuseInformation (e.g. data definitions, experimental parameters) or dataQualityInformation. The type of related object
should therefore be left ambiguous by the automated mapping process.
Mapping from EPrints to RIF-CS

The following table provides a mapping to populate a RIF-CS Collection record from an EPrints record. This mapping takes account of variations to the standard EPrints metadata profile made by the ReCollect plugin (developed by the University of Essex) and the University of Glasgow.¹

<table>
<thead>
<tr>
<th>RIF-CS 1.5 element</th>
<th>Source (using EPrints record)</th>
</tr>
</thead>
<tbody>
<tr>
<td>collection[dateAccessioned]</td>
<td>datestamp</td>
</tr>
<tr>
<td>identifier[type=local]</td>
<td>id_number</td>
</tr>
<tr>
<td>name[type=primary] &gt; namePart</td>
<td>title</td>
</tr>
<tr>
<td>name[type=alternative] &gt; namePart</td>
<td>alt_title</td>
</tr>
<tr>
<td>dates[type=dc.available] &gt; date[type=dateFrom]</td>
<td>date_embargo</td>
</tr>
<tr>
<td>dates[type=dc.created] &gt; date</td>
<td>collection_date (ReCollect)</td>
</tr>
<tr>
<td>dates[type=dc.dateSubmitted] &gt; date</td>
<td>datetstamp</td>
</tr>
<tr>
<td>dates[type=dc.issued] &gt; date[type=dateFrom]</td>
<td>revision</td>
</tr>
<tr>
<td>location &gt; address &gt; electronic[type=url] &gt; value</td>
<td>Derived from id_number</td>
</tr>
<tr>
<td>subject[type=local]</td>
<td>keywords</td>
</tr>
<tr>
<td>subject</td>
<td>subjects</td>
</tr>
<tr>
<td>description[type=full]</td>
<td>abstract</td>
</tr>
<tr>
<td>description[type=lineage]</td>
<td>provenance (ReCollect)</td>
</tr>
<tr>
<td>description[type=note]</td>
<td>note</td>
</tr>
<tr>
<td>coverage &gt; temporal &gt; date</td>
<td>temporal_cover (ReCollect)</td>
</tr>
<tr>
<td>coverage &gt; spatial[type=text]</td>
<td>geographic_cover (ReCollect)</td>
</tr>
<tr>
<td>coverage &gt; spatial[type=iso19139dcmiBox]</td>
<td>bounding_box (ReCollect)</td>
</tr>
<tr>
<td>relatedInfo[type=collection] &gt; identifier</td>
<td>related_resources (ReCollect)</td>
</tr>
<tr>
<td>relatedInfo[type=collection] &gt; relation[type]</td>
<td>related_resources &gt; relationType (ReCollect)</td>
</tr>
<tr>
<td>rights &gt; accessRights</td>
<td>security, restrictions (Glasgow), accessLimitations (Glasgow)</td>
</tr>
<tr>
<td>rights &gt; licence</td>
<td>license</td>
</tr>
<tr>
<td>citationInfo &gt; citationMetadata &gt; identifier</td>
<td>id_number</td>
</tr>
</tbody>
</table>

1. The metadata fields used by UK EPrints users to describe datasets is being tracked at http://tinyurl.com/or34774.
7.1 Related Objects

- The holding repository is represented in the Registry as a Collection of type ‘repository’. Each dataset object includes the repository object as a Related Object with Relation type ‘isLocatedIn’. The repository object lists each dataset object as a Related Object with Relation type ‘isLocationFor’.

- Individuals and corporate entities contributing to the dataset are represented in the Registry as Parties. The dataset object lists each creator as a Related Object with Relation type ‘hasPrincipalInvestigator’. Each creator object includes the dataset object as a Related Object with Relation type ‘isPrincipalInvestigatorOf’.

- The project that produced the dataset (recorded in the ReCollect grant element or EPrints projects element) would be represented by a minimal Activity object:

```plaintext
Sample Record 1

registryObject[group=...]
  key=...
  originatingSource=...
  activity[type=project, dateModified=...]
    identifier=(grant number)
    relatedObject
      key=(Collection key)
      relation[type=hasOutput]
      relatedObject
        key=(Party key)
        relation[type=isFundedBy]
```

The dataset object would have a reciprocal relation of type ‘isOutputOf’.

- Funding information (recorded in ReCollect records using the funders element) would be represented by a Party object, related to the dataset object via the above project (Activity) object:
Theregistryshouldalreadybepopulatedwithmajorfundingbodies,soinmost
instancesthiswillinvolvemakinginformationtotheexistingrecordratherthan
creatinganewone.
Mapping from OAI-PMH Dublin Core to RIF-CS

The following table provides a mapping to populate a RIF-CS Collection record from an OAI-PMH Dublin Core (oai_dc) record.¹

<table>
<thead>
<tr>
<th>RIF-CS 1.5 element</th>
<th>Source (using oai_dc record)</th>
</tr>
</thead>
<tbody>
<tr>
<td>identifier²</td>
<td>dc:identifier</td>
</tr>
<tr>
<td>name &gt; namePart</td>
<td>dc:title</td>
</tr>
<tr>
<td>dates[type=dc.issued] &gt; date[type=dateFrom]</td>
<td>dc:date</td>
</tr>
<tr>
<td>location &gt; address &gt; electronic[type=url] &gt; value</td>
<td>dc:identifier (if URL)</td>
</tr>
<tr>
<td>subject[type=local]</td>
<td>dc:subject</td>
</tr>
<tr>
<td>description[type=full]</td>
<td>dc:description</td>
</tr>
<tr>
<td>coverage &gt; temporal &gt; date³</td>
<td>dc:coverage (if parsed as date information)</td>
</tr>
<tr>
<td>coverage &gt; spatial[type=text]</td>
<td>dc:coverage (if not parsed as date information)</td>
</tr>
<tr>
<td>relatedInfo &gt; identifier²</td>
<td>dc:relation</td>
</tr>
<tr>
<td>relatedInfo &gt; relation[type=hasAssociationWith] &gt; description</td>
<td>‘Unknown’</td>
</tr>
<tr>
<td>rights &gt; rightsStatement</td>
<td>dc:rights</td>
</tr>
<tr>
<td>citationInfo &gt; citationMetadata &gt; identifier²</td>
<td>dc:identifier</td>
</tr>
<tr>
<td>citationInfo &gt; citationMetadata &gt; contributor &gt; namePart</td>
<td>dc:creator, dc:contributor</td>
</tr>
<tr>
<td>citationInfo &gt; citationMetadata &gt; title</td>
<td>dc:title</td>
</tr>
<tr>
<td>citationInfo &gt; citationMetadata &gt; publisher</td>
<td>dc:publisher</td>
</tr>
<tr>
<td>date[type=publicationDate, available, issued]</td>
<td>dc:date</td>
</tr>
</tbody>
</table>


2. The type of identifier would be inferred from pattern matching, falling back to ‘local’ if no match is found.

3. A single date value in the oai_dc record would be represented by date[type=dateFrom] in the RIF-CS record. A date range in the oai_dc record, such as ‘2004-03-02/2005-06-02’, would be represented by date[type=dateFrom] (‘2004-03-02’) and date[type=dateTo] (‘2005-06-02’) in the RIF-CS record.
8.1 Related Objects

• Individuals and corporate entities contributing to the dataset are represented in the Registry as Parties. The dataset object lists each creator as a Related Object with Relation type ‘hasPrincipalInvestigator’. Each creator object includes the dataset object as a Related Object with Relation type ‘isPrincipalInvestigatorOf’.

• If the value of dc:source matches an object already in the Registry, a Relation of type ‘isDerivedFrom’ is added to the current record (indicating the referenced object) and a Relation of type ‘hasDerivedCollection’ is added to the referenced record (indicating the current object).
9.1 High priority metadata from harvested records

While RIF-CS has little in the way of mandatory metadata elements, a certain level of information will be needed if a harvested record is going to make a useful contribution to the Registry. The two highest priorities are

1. ensuring enough information can be displayed about the dataset so that users of the Registry can determine if a dataset might be useful for them;
2. gathering enough information to allow the dataset to show up in search results and the browsing interfaces.

The following items of information are listed in approximate order of importance, with the most important first. Shown under each heading are the relevant RIF-CS element(s), an indication of how the information is used by the Registry, and whence the information is harvested.

9.1.1 Dataset name

RIF-CS

name > namePart\[TYPE=primary\],
citationInfo > citationMetadata > title

Usage
Used as the title of the record, in lists of records, and in the sample citation.

DDI
codeBook > stdyDscr > citation > titlStmt > titl

NERC
1 – Title

DataCite
Title

EPrints
title

oai_dc
dc:title

9.1.2 Full description

RIF-CS
description\[TYPE=full\]

Usage
Used prominently in the displayed record, and in free text searches.

DDI
codeBook > stdyDscr > stdyInfo > abstract

NERC
4 – Abstract

DataCite
Description[DESCRIPTIONTYPE=Abstract]

EPrints
abstract

oai_dc
dc:description
9.1.3 Identifier

**RIF-CS**
identifier,
citationInfo > citationMetadata > identifier

**Usage**
Used prominently in the displayed record, in free text searches, and in the sample citation.

**DDI**
`'10.5255/UKDA-SN-' + UKDA ID + '-1'`

**NERC**
36 – Unique resource identifier > Code

**DataCite**
Identifier[IDENTIFIERTYPE=DOI]

**EPrints**
id_number

**oai_dc**
dc:identifier

9.1.4 Subject

**RIF-CS**
subject

**Usage**
Used prominently in the displayed record, in free text searches, and as a browsing hierarchy.

**DDI**
codeBook > stdyDscr > stdy > subject > keyword[VOCAB=S],
codeBook > stdyDscr > stdy > subject > topClas

**NERC**
5 – Topic category,
6 – Keyword > Keyword value

**DataCite**
Subject

**EPrints**
keywords,
subjects

**oai_dc**
dc:subject

9.1.5 URL

**RIF-CS**
location > address > electronic[TYPE=url] > value,
citationInfo > citationMetadata > url

**Usage**
Used in the displayed record to provide access, and in the sample citation.

**DDI**
`http://dx.doi.org/10.5255/UKDA-SN-' + UKDA ID + '-1'

**NERC**
36 – Unique resource identifier if resolvable, otherwise 19 – Resource locator > Resource locator URL

**DataCite**
`http://dx.doi.org/'' + Identifier[IDENTIFIERTYPE=DOI]

**EPrints**
derived from id_number

**oai_dc**
dc:identifier (if URL)

9.1.6 Date published, issued, made available

**RIF-CS**
dates[TYPE=dc.issued] > date[TYPE=dateFrom],
citationInfo > citationMetadata > date[TYPE=publicationDate, available, issued]

**Usage**
Used as a search refinement parameter, and in the sample citation.

**DDI**
codeBook > stdyDscr > citation > distStmt > distDate
NERC 8 – Dataset reference date > date type = publication, date
DataCite PublicationYear,
Date[DATETYPE=Available, Issued]
EPrints published, issued: revision,
available: date_embargo
oai_dc dc:date

9.1.7 Creator

RIF-CS citationInfo > citationMetadata > contributor > namePart
Usage Used in the sample citation, and to connect to Party records.
DDI codeBook > stdyDscr > citation > rspStmt > AuthEntry
NERC 23 – Responsible organisation > Individual name
DataCite Creator[CREATORNAME]
EPrints creators
oai_dc dc:creator,
dc:contributor

9.1.8 Rights information

RIF-CS rights > …
Usage Used in the displayed record.
DDI rightsStatement: codeBook > stdyDscr > citation > prodStmt > copyright,
accessRights: codeBook > stdyDscr > dataAccs > useStmt > restrctn, conditions
NERC accessRights: 25 – Limitations on Public Access
26 – Use Constraints
DataCite rightsStatement: Rights
EPrints accessRights: security, restrictions (Glasgow), accessLimitations (Glasgow),
licence: license
oai_dc rightsStatement: dc:rights

9.1.9 Spatial Coverage

RIF-CS coverage > spatial
Usage Used in the displayed record, and as a browsing tool.
DDI codeBook > stdyDscr > stdyInfo > sumDscr > geogCover, geogUnit, nation,
codeBook > stdyDscr > stdyInfo > subject > keyword[VOCAB=G]
NERC 44 – Bounding box,
5 – Extent > Extent name
DataCite GeoLocation
EPrints bounding_box,
geographic_cover
oai_dc dc:coverage (if not parsed as date information)
9.1.10  **Publisher**

**RIF-CS**  \[citationInfo > citationMetadata > publisher\]

**Usage**  Used in the sample citation.

**DDI**  \[codeBook > stdyDscr > citation > distStmt > distrbtr\]

**NERC**  23 – Responsible organisation > Organisation name

**DataCite**  Publisher

**EPrints**  publisher

**oai_dc**  dc:publisher

DATACITE METADATA WORKING GROUP, *DataCite Metadata Schema for the Publication and Citation of Research Data*, version 3.0 (DataCite Consortium, 2013), DOI: 10.5438/6808.


